

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 16, 2003, 15:50:54 ; Search time 2492.14 Seconds
(without alignments)
16465.781 Million cell updates/sec
Title: US-09-497-967-102
Perfect score: 1410
Sequence: 1 atgaagaacacactctggt.....cttactacctgctgtaataa 1410

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	784.4	55.6	3026	3	AF324424	AF324424 Ichthyoph
2	63.2	4.5	2486	3	AF140273	AF140273 Ichthyoph
3	57.2	4.1	1520	3	AF405431	AF405431 Ichthyoph
4	55.2	3.9	125020	9	AF429315	AF429315 Homo sapi
5	52.2	3.7	1249	3	ICVIMANT	M92907 Ichthyophh
6	50	3.5	125020	9	AF429315	AF429315 Homo sapi
c	43.4	3.1	230372	2	AC073693	AC073693 Mus muscu
7	41.4	2.9	159764	9	AC093866	AC093866 Homo sapi
8	41	2.9	643	8	AF374493	AF374493 Thalassio
9	41	2.9	179217	2	AC099748	AC099748 Bos tauru
10	39.4	2.8	591	8	AF374552	AF374552 Thalassio
11	38.6	2.7	643	8	AF374492	AF374492 Thalassio
12	38.6	2.7	16814	9	HSCOLPDGF	X98705 H.sapiens D
c	38.6	2.7	341950	1	AP003600	AP003600 Nosloc sp
13	38.4	2.7	224271	9	AC026803	AC026803 Homo sapi
c	38.2	2.7	15294	1	AE006936	AE006936 Mycobacte
14	38.2	2.7	31300	1	MTV035	AE021930 Mycobacte
15	38.2	2.7	152777	2	AC125855	AC125855 Rattus no
c	38.2	2.7	233345	4	AJ421481	AJ421481 Bos tauru
16	38.2	2.7	338150	1	AP005278	AP005278 Corynebac
17	38.2	2.7	349980	6	AX127147	AX127147 Sequence
18	37.8	2.7	559	8	AF374543	AF374543 Thalassio
19	37.8	2.7	591	8	AF374540	AF374540 Thalassio
20	37.8	2.7	591	8	AF374542	AF374542 Thalassio
21	37.8	2.7	591	8	AF374544	AF374544 Thalassio
22	37.8	2.7	591	8	AF374545	AF374545 Thalassio
23	37.8	2.7	591	8	AF374547	AF374547 Thalassio
24	37.8	2.7	591	8	AF374548	AF374548 Thalassio
25	37.8	2.7	591	8	AF374549	AF374549 Thalassio
26	37.8	2.7	591	8	AF374550	AF374550 Thalassio
27	37.8	2.7	591	8	AF374551	AF374551 Thalassio
28	37.8	2.7	643	8	AF374490	AF374490 Thalassio
29	37.8	2.7	643	8	AF374491	AF374491 Thalassio
30	37.8	2.7	643	8	AF374494	AF374494 Thalassio
31	37.8	2.7	643	8	AF374495	AF374495 Thalassio
32	37.8	2.7	643	8	AF374496	AF374496 Thalassio
33	37.8	2.7	643	8	AF374497	AF374497 Thalassio
34	37.8	2.7	643	8	AF374498	AF374498 Thalassio
35	37.8	2.7	643	8	AF374499	AF374499 Thalassio
36	37.8	2.7	643	8	AF374500	AF374500 Thalassio
37	37.8	2.7	643	8	AF374501	AF374501 Thalassio
38	37.8	2.7	643	8	AF374502	AF374502 Thalassio
39	37.8	2.7	643	8	AF374503	AF374503 Thalassio
40	37.8	2.7	643	8	AF374504	AF374504 Thalassio
41	37.8	2.7	643	8	AF374505	AF374505 Thalassio

ALIGNMENTS

RESULT 1
AF324424
LOCUS AF324424 3026 bp DNA linear INV 27-FEB-2002
DEFINITION Ichthyophthirius multifiliis immobilization antigen isoform
(IAG52A) gene, complete cds.
ACCESSION AF324424
VERSION AF324424.1 GI:12698726
KEYWORDS
SOURCE Ichthyophthirius multifiliis.
ORGANISM Ichthyophthirius multifiliis.
REFERENCE 1 (bases 1 to 3026)
AUTHORS Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Ophryoglenina; Ichthyophthirius.
Lin, Y., Lin, T.L., Wang, C.C., Wang, X., Stieger, K., Klopfleisch, R.
and Clark, T.G.

TITLE Variation in primary sequence and tandem repeat copy number among
1-antigens of Ichthyophthirius multifiliis
JOURNAL Mol. Biochem. Parasitol. 120 (1), 93-106 (2002)
MEDLINE 21839613
PUBMED 11849709
REFERENCE 2 (bases 1 to 3026)
AUTHORS Lin, Y., Lin, T.-L. and Clark, T. G.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2000) Microbiology and Immunology, NYSCVM Cornell
University, Ithaca, NY 14853, USA
FEATURES
source Location/Qualifiers
1..3026
/organism="Ichthyophthirius multifiliis"
/macronuclear
/serotype="D"
/isolate="G5"
/db_xref="taxon:5932"
1285..2817
/gene="IAG52A"
1285..2817
/gene="IAG52A"
/product="Immobilization antigen isoform"
1330..2736
/gene="IAG52A"
/note="52 kDa protein"
/codon_start=1
/transl_table=6
/product="Immobilization antigen isoform"
/protein_id="AAK01661.1"
/db_xref="GI:12698727"
/translation="MKNILVILIISLFINIKSNCPVGTETNATAGVDDLGTPANCVNCRFYNNAAAEVPGASTCPQPKKDGAGQNPATANLVTCNVKPCPAGTAIGCATDYAAITCEVNCRIINFENAPNEAGASTACPVNVRVGGALTAGNAATIVAGCNVACPTGALDDGYTQVRSFTECVKRLNIFYNGNNGNTPNPGKSOCTPCPAIKPANVAQLGNDATLTAOCNVACPDGTLSAAGVNNVAQNTFECTNCAENFYNNAPNENFNCSTCLPDPANKDYGAEATAGGARTLAKQCNACPDGTIAGSATNIVYLQTECLNCAANFYDGNNGFAGSRRCKACPAKQVAVATAGGTATLIAQCALECPAGTVLTDGTTSTYKQAAECGKCAANFYTKQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNIQDFANFLSLSLLLISYLL"
BASE COUNT 1177 a 316 c 355 g 1178 t
ORIGIN
Query Match 55.6%; Score 784.4; DB 3; Length 3026;
Best Local Similarity 72.3%; Pred. No. 3.8e-220;
Matches 1019; Conservative 0; Mismatches 391; Indels 0; Gaps 0;
QY 1 ATGAAGAACAACTCTGGTGATCATCTCTCTGTCATCAACACAGATCAAGTCT 60
DB 1330 ATGAATAATATATTTAGTAATATTGATTATTTATTTATCAATTAATTAATCT 1389
QY 61 GCTAACTGCTCTGGGAACCGAGACCAACACCGCTGGACAGGTGGACACCTGGGAACC 120
DB 1390 GCTAATTTGCTCTGGAACTGAACTAACACAGCGGATAAGTTGATGATCTAGGAAC 1449
QY 121 CTTGCTAACTGTGAAGCTGACAGAACTTCTACTACAAACACGCTGCTGCTTTCGTG 180
DB 1450 CTTGCAAAATGTTAATTTAGTAAAGAACTTTTATTAATAAATGCTGCTGCTTTCGT 1509
QY 181 CTTGGAGCTTCTACCTGTACCCCTTGTCTCAGAAGAAGACGCTGAGCTCAGCCCTAAC 240
DB 1510 CTTGGTGCTAGTACGTGACACCTTGTCCATAAAAAAGATGCTGGTGCCTTAACCAAT 1569
QY 241 CCTCCTGCTACCGCTAACCTGGTGACCCAGTGTAACTGAAAGTGTCTGCTGGAACCGCT 300
DB 1570 CCACCTGCTACTGCTAATTTAGTACATAATGTAACGTTTAAATGCCCTGCTGGTACCGCA 1629
QY 301 ATCGCTGGAGAGCTACCGACTACGCTGTATCATCACCAGAGTGTGTGAAGTGTCCATC 360
DB 1630 ATTCAGGTGGAGCAACAGATTATGCAGCAATAATACACAGAATGTGTTAATTGTAGAAAT 1689
QY 361 AACTTCTACAAACGAGCGCTCTTAATCTCAACGCTGGAGCTTCTACCTGTACCGCTTCT 420
DB 1690 AATTTTATAATGAATAATGCTCCAAATTTTAATGCAGGTGCTAGTACATGCACAGCTTCT 1749

QY 421 CCTGTGAACCCGCTGGGAGGAGCTTCTGACCGCTGGAACAGCTGTACCATCTGGTGGCTCAG 480
DB 1750 CCGGTAAACAGAGTGGTGGTGCATTGACTCTGGTGAATGCGCGTACCATAGTCCGATAA 1809
QY 481 TGTAACTGGCTTGTCTTACCGGAACCGCTCTGACGACGAGGAGTACCACGACTACGCTG 540
DB 1810 TGTAACTGGCTTGTCTTACCGGAACCGCTCTGACGACGAGGAGTACCACGACTACGCTG 1869
QY 541 CGCTCTTTTCAACCGAGTGTGAAGTGTGCGCTGAACCTTCTACTACAAAGGAAACACGGA 600
DB 1870 AGATCATTTACAGAAATGTTAAATGTAGACTTAACCTTTACTATAATGTTAAATAGGT 1929
QY 601 AACACCCCTTTCAACCGCTGGAAAGTCTCAGTGTACCCCTTGTCTGTATCAACCCCTGCT 660
DB 1930 AATACTCCTTTCAATCCAGGTAAAAGTTAATGCACACCTTGTCCGGCAATTTAAACCTGCT 1989
QY 661 AACGTGGCTCAGGCTACCCCTGGGAACGACGCTACCAATCACCGCTCAGTGTAAACGCTGCT 720
DB 1990 AATGTTGCTTAAAGCTACTTTAGGTAATGATGCTACAATAACCGCATATGTTAAACGTTGCA 2049
QY 721 TGTCTGACGGAACCATCTCTGCTGGAGTGAACAACTGGTGGTTCAGAACACCGAG 780
DB 2050 TGCCCTGATGCTACTATAAGTCTGCTGGAGTAAATAATTTGGGTAGCACAAACACATGAA 2109
QY 781 TGTACCAACTGTGCTCTTAACCTTCTACAAACAAACGCTCTCTAACTTCAACCCCTGGAAAC 840
DB 2110 TGTACTAATTTGCTGCTCTTAACCTTTTACAAATAATAATGCTCTCTAAATTTCAATCCAGTAAT 2169
QY 841 TCTACTGCTGCTGCTCTGCTCTTAAACAAAGACTACGAGCTACCGCTACCGCTGGAGGA 900
DB 2170 AGTACACGCTTACCTTGCCAGCAAAATAAGATTATGTTGCTGAAGCCACTGCGAGGTGGT 2229
QY 901 GCTGTACCCCTGGCTTAAGCAGTGAACATCGCTTGTCTGTACGGAACCGCTATCGCTTCT 960
DB 2230 GCCGCTACTTTAGCCAAATATGTAATATTCATGCCCTGATGTTGCTACTGCAATTTGCTAGT 2289
QY 961 GGAGCTACCAACTAGCTGATCCTCGACACCGAGTGTCTGAACTGTGCTGCTTAACCTTCTAC 1020
DB 2290 GGAGCAACTAATATGTAATATTATAAACAAGATGCTCTAAATTTGCTGCTTAACCTTTAT 2349
QY 1021 TTCGACGGAACAACTTCCAGCTGGATCTTCTGCTGTAAAGCTTGTCTGCTTAACAAG 1080
DB 2350 TTTGATGTAATAATTTCTAGCGAGAGTAGTAGTACCAAGCATGTTCCAGCAAAATAA 2409
QY 1081 GTGACGAGGAGTGTGGCTTACCGCTGGAGAACCGCTACCCCTGATCGCTCAGTGTGCTGTG 1140
DB 2410 GTTTAAGCGCTGTAGCAACTGCAGGTGTTACTGCTACTTTTAATGTCATAATGTGCCCTT 2469
QY 1141 GAGTGTCTGCTGGAACCGTCTGACCGACGGAACCCACTCTACCTACAAACGAGCTGCT 1200
DB 2470 GAATGCCCTGCTGTACTGTACTCACCAGTGAACAACATCTACTTTAAATAAGCAGCA 2529
QY 1201 TCTGAGTGTGTAAGTGTGCTGAACCTTCTACACACCAAGCAGACCGACTGGGTGGCT 1260
DB 2530 TCTGAATGTGTTAAATGTGCTGCCAACHTTTATCTACAAAATAAAGTGGTAGCA 2589
QY 1261 GGAATCGACACTGTACCTCTTTGTAACAAGAGCTGACCTCTGGAGCTGAGGCTAAACCTG 1320
DB 2590 GGTATTGATACATGTACTAGTTGTTAATAAAAAATTAACCTTCTGGCGCTGAAGCTAAATTA 2649
QY 1321 CTTGAGTGTCTGAAGAACAATCCAGTGTGACTTCGCTTAACCTTCTCTCTCTCTCTG 1380
DB 2650 CTTGAATCTGCTAAAAAAAATAATAATGTTGATTTGCTGCTAAATTTTATCAATTTCCCTTA 2709
QY 1381 CTGCTGCTCTCTTACTACCTGCTGTAATAA 1410
DB 2710 TTATTTGATTTCTTATTATTATTATGATGA 2739

RESULT 2
AF140273
LOCUS AF140273 2486 bp DNA linear INV 19-MAY-1999

```

DEFINITION Ichthyophthirius multifiliis immobilization antigen precursor
              (IAG48) gene, complete cds.
ACCESSION AF140273
VERSION AF140273.1 GI:4868370
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 2486)
AUTHORS Clark,T.G., Lin,T.L., Jackwood,D.A., Sherrill,J., Lin,Y. and
          Dickerson,H.W.
TITLE The gene for an abundant parasite coat protein predicts tandemly
          repetitive metal binding domains
JOURNAL Gene 229 (1-2), 91-100 (1999)
MEDLINE 99196987
PUBMED 10095108
REFERENCE 2 (bases 1 to 2486)
AUTHORS Gaertig,J., Gao,Y., Tishgarten,T., Clark,T.G. and Dickerson,H.W.
TITLE Surface display of a parasite antigen in the ciliate Tetrahymena
          thermophila
JOURNAL Nat. Biotechnol. (1999) In press
REFERENCE 3 (bases 1 to 2486)
AUTHORS Clark,T.G., Lin,T.-L., Jackwood,D.A. and Dickerson,H.W.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-1999) Microbiology & Immunology, Cornell
          University, College of Veterinary Medicine, Ithaca, NY 14853, USA
FEATURES
    source
        1..2486
            /organism="Ichthyophthirius multifiliis"
            /serotype="A"
            /isolate="G1"
            /db_xref="taxon:5932"
            399..1845
            /gene="IAG48"
            399..1845
            /gene="IAG48"
            /product="immobilization antigen precursor"
            433..1761
            /gene="IAG48"
            /note="surface protein; 48 kDa i-antigen"
            /codon_start=1
            /transl_table=6
            /product="immobilization antigen precursor"
            /protein_id="AAD31283.1"
            /db_xref="GI:4868371"
            /translation="MKYNLLILILISLFINELRAVPCPDGTQTQAGLTDVGAADLGTC
            VNCRPNEYNGAAGANGNQPAANNAARGICVPCQINRVGVTNAGDLATLQTC
            STQCTGTALDDGVTDFRSAACQVKCPNFYNGSPQGEAFGVQVFAAGAAAGV
            AAVTSQCVPCOLNKNDSPATGAQAANLQCSNQCPTGLDDGVTLVFNTSATLCVK
            CRPNFYNGSPQGEAFGVQVFAAGAAAGVAAVTSQCVPCQINKNDSPATGAQAANL
            ATQCSQCTGTATQDGTVLVFSNSQCSQCIANYFFENGFEAGKSOCLKCPVSKTT
            PAHAPGNTATQCLTCTCPAGTVLDGTSNFEVASATECTKSAGFFASKTGTAG
            TDTCTCKKLTSATAKVYAEATQKVQCASTTFPAKFLISLILFISFYLL"
            433..492
            /gene="IAG48"
            493..1758
            /gene="IAG48"
            /product="immobilization antigen"
            896 a 310 c 321 g 959 t

    sig_peptide
    mat_peptide

BASE COUNT 896 a 310 c 321 g 959 t
ORIGIN

Query Match 4.5%; Score 63.2; DB 3; Length 2486;
Best Local Similarity 52.2%; Pred. No. 4.5e-07;
Matches 140; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 1098 TACCGCTGAGAACCGCTACCCCTGATCGCTCAGTGTGCTCTGGAGTGTCTCTGCTGGAAC 1157
Db 1443 TGCTCCAGGTAATCTGCTACTTAAGCCACATAATGTTTGACCAACATGCTCTGCTGTAC 1502

QY 1158 CGTGTGACCGACGGAACCGCTCTTACCTACAACGAGCGCTCTTCTGAGTGTGTAAGTG 1217
Db 1503 AGTACTTGATGATGACATCACTAATTTTGTAGTTCGCGCAACTGAATGCTAAATG 1562

```

```

QY 1218 TGCTGCTAACTTTTACACCACCAAGCAGACACCGACTGGTGGCTGAATCGACACCTGTAC 1277
Db 1563 TTCTGCTGGCTTTTGTGCATCAAAAACAACACTGGTTTTACAGCAGGTACTGATACATGTAC 1622

QY 1278 CTTCTGTACAAGAAGCTGACCTCTGGAGCTGAGCTAACCTGCTGAGTGTCTGTGAAGA 1337
Db 1623 TGAATGTACTAAAAAATAACTTCTGGTCCACAGCTAAAGTATATGCTGAAGCTACTCA 1682

QY 1338 GAACATCCAGTGTGACTTCGGCTAACTTC 1365
Db 1683 AAAAGTATAATGCGCTCCACTACTTTC 1710

RESULT 3
AF405431
LOCUS
DEFINITION Ichthyophthirius multifiliis 52kDa immobilization antigen variant B
          protein mRNA, complete cds.
ACCESSION AF405431
VERSION AF405431.1 GI:15290741
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1520)
AUTHORS Lin,Y., Wang,J.C. and Clark,T.C.
TITLE Variation in primary sequence and tandem repeat copy number among
          i-antigens of Ichthyophthirius multifiliis
JOURNAL Mol. Biochem. Parasitol. 120 (1), 93-106 (2002)
MEDLINE 21839613
PUBMED 11849709
REFERENCE 2 (bases 1 to 1520)
AUTHORS Lin,Y., Wang,J.C. and Clark,T.C.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2001) Microbiology and Immunology, NYSCVM Cornell
          University, Ithaca, NY 14853, USA
FEATURES
    source
        1..1520
            /organism="Ichthyophthirius multifiliis"
            /macronuclear
            /serotype="D"
            /isolate="G5"
            /db_xref="taxon:5932"
            41..1423
            /note="IAG52B[G5]"
            /codon_start=1
            /transl_table=6
            /product="52kDa immobilization antigen variant B protein"
            /protein_id="AAK94941.1"
            /db_xref="GI:15290742"
            /translation="MKFNILILILISLFINELRAVNCNPAANGOSDPTGAADINTC
            THCKHFYFNGGNPAGAPGVQSPQCIACQVHKADSHRGGDANLAQGSNL
            CPAGTAVEDGDSPTFTQSLQCVNCPNFYFNGNPTGQAPGAGDPTQLIANPDLAN
            NPEVPNVSSPGQCVACQVKNQSDSLRPAQANLATQCNEGTDTATQDGAIFTYQ
            SISQCTFCKVDYFNGNPNQNGQFTPGQLIANPDATTAAGTIPVPGPNKQV
            CESKNTNSRSGLEANLAQCGTECPAGTLVDGVTPTVTVSLQCVNCKAGFYQNS
            NFEAGKSCNKCASVSKTGSASVPGNSATSATOCNDCPAGTVDDGTSTNFFVALSEC
            TKQANFYASKTSGFAAGTDTCTCECKSLTSATAKVYAEATQKACASSTFPAKFLSM
            SLIFISFYLL"
            493 a 255 c 256 g 516 t

BASE COUNT 493 a 255 c 256 g 516 t
ORIGIN

Query Match 4.1%; Score 57.2; DB 3; Length 1520;
Best Local Similarity 52.0%; Pred. No. 2.5e-05;
Matches 159; Conservative 0; Mismatches 138; Indels 9; Gaps 1;

QY 1114 GCTACCTGATCCGTGAGTGTGCTGGAGTGTCTCTGGAAGTGTCTGACCGACGGA 1173
Db 1121 GCTACTTCAGCCACATAATGTTAAAACGATTGCGCTGCTGGTACATGTTGATGATGCT 1180

```

```
QY 1174 ACCACCTCTACCTACCAAGCAGCTCTCTCTGAGTGTGTGAAGTGTGCTGCTAACTTCTAC 1233
|| ||| |
Db 1181 ACATCAACTAATTTTGTAGCTTTAGCAAGTGAATGCTACTAAATGTTAGGCTAACTTTAT 1240
|| ||| |
QY 1234 ACCACCAAGCAGCAGCTGGTGGCTGGAATCGACACCTGTACCTCTGTGTAAACAAGAG 1293
|| ||| |
Db 1241 GCATCAAAACATCTGGTTTTCGACAGGTACTGATACATGTACGATGTCTTAAAAAAA 1300
|| ||| |
QY 1294 CTGACCTCTGGAGCTGAGCTAACTGCTGAGTCTGCTGAAGAACAATCCAGTGTGAC 1353
|| ||| |
Db 1301 TTAACCTCTGGTGCTACAGCTAACTATATGCTGAAGCTACTTAAAAAGCATAATGCGCC 1360
|| ||| |
QY 1354 -----TTGCGCTAACTCTGCTATCTCTCTGCTGCTGCTGCTTACTACCTGCTG 1404
|| ||| |
Db 1361 AGTTCACCTTTCGCAAAATTTTATCAATGCTCTTAATATTATTTCTTCTATTGTTG 1420
|| ||| |
QY 1405 TAATAA 1410
|| ||| |
Db 1421 TGAAGA 1426
|| ||| |

RESULT 4
AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 125020)
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
21583737
11694876
2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit=ctg
complement(<36507..>36887)
/gene="JPH3"
/note="JPH3"
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translaton="MSSGGRFNPDGSGYCGGWEDGKAHGHGVCYTPKGGQGYTSGWS
HGFEVLGVYTPSGNTYQGTWAGQKRHGIGLESKSKWYKGEWTHGFKRGYGVRECA
```

```
NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN
Query Match 3.9%; Score 55.2; DB 9; Length 125020;
Best Local Similarity 10.9%; Pred. No. 0.00014;
Matches 108; Conservative 421; Mismatches 452; Indels 6; Gaps 3;
QY 199 ACCCTTGTCTCAGAGAAGGAGCGTGGAGCTCAGCTAACCTCTCTGCTACCGCTAAC 258
|| ||| |
Db 16868 ATCCARKKMKSHAGRRRCYYSKSWMBMSVSYSVKHSHASBSCHMBKCHTWS 16927
|| ||| |
QY 259 CTGCTGACCCAGTGAAGTGAAGTCTCTCTGCTGGAACCGCTATCGCTGGAGGAGTACC 318
|| ||| |
Db 16928 CMSMMYKSSWGWSSWGCWRRRSKWKWYSRGMSRSKSMRYTGGSKMMRSMNCTSSC 16987
|| ||| |
QY 319 GACTACGCTGTATCATCACCAGGTGTGTAAGTGTGCGATCAACTTCTACAACGAGAAC 378
|| ||| |
Db 16988 YASMCMMCCWSCCSCCMRSCCRSYCCCMRYCCACKCYMSYWTMSASYSRKRYSYWRSM 17047
|| ||| |
QY 379 GCTCTTAACCTCAACGCTGGAGCTTCTACCTGTACCGCTTGTCTGTGAACCGCGTGGGA 438
|| ||| |
Db 17048 KCWS--RMCSSRSKCKSRCCGSGMGKGGKSGYGRKTRKSRGKMGAKWYMYRRSM 17105
|| ||| |
QY 439 GGAGCTCTGACCGCT---GGAACGCTGTACCATCTGCTGCTGCTAGTGTACGCTGGCTGT 495
|| ||| |
Db 17106 RKMYSKSGMYCMYCWGRRGCGYCSCTSRSAMCCSYCYAKCKMCMYCYCYGMSMKGY 17165
|| ||| |
QY 496 CTTACCGGAACCGCTCTGGACGAGGAGTACCACCGCTACGCTGCGCTCTTTCACCGAG 555
|| ||| |
Db 17166 YYACSYGMSMSKSKYMRGTYSTSGGCCCTTTTCCCCCNANTGGGAAGCTTTTNCNK 17225
|| ||| |
QY 556 TGTGTGAAGTGTGCTGGAAGTCTTCTACTACAAACGGAACAAACGGAACACACCCCTTCAAC 615
|| ||| |
Db 17226 TYSYRKRNGCAMCKYNNYNNWSRSCRAGMSCTKYKSSMTMSWASYCWSMYSYCSMR 17285
|| ||| |
QY 616 CTTGAAAGTCTCAGTGTACCCCTTGCTCTCTCTGCTATCAAGCTGTCTGCTGCTGCTGCTGCT 675
|| ||| |
Db 17286 SMASRGSWSYMKMKMSWRMSYMKCCMKCMYCMRSMRSMRSGMSYMYWASWKSRR 17345
|| ||| |
QY 676 ACCCTGGGAACGAGCTACCATCACCGCTCAGTGTAAACGCTGCTGCTGCTGCTGCTGCTGCT 735
|| ||| |
Db 17346 GCYCTRCYWMSSSKSKYSSYMMRSMRSMRSGMSYMYWASWKSRR 17405
|| ||| |
QY 736 ATCTCTGCTGTGAGTGAACAACACTGGGTGGCTCAGACACCGAGTGTACCACTGCTGCT 795
|| ||| |
Db 17406 MMSSRCYRCAGCMKMGYMYRMCWNRWGRWGRWGRWGRWGRWGRWGRWGRWGRWGRWGRWGR 17465
|| ||| |
QY 796 CTTAACTTCTACAACAACGCTCTCTAACTTCAAC--CTGGAACACTCTACCTGCTGCTGCC 854
|| ||| |
Db 17466 MSRCMMMSKSYCASCRCSCAMSGTYMKCASYNCMSYWCCTWCCTSYMYRCTCCKGWS 17525
|| ||| |
QY 855 TTGCTCTGCTAAACAGGACTACGAGCTGAGGCTACCGCTGAGGAGTGTCTACCTGCTGCTGCTGCT 914
|| ||| |
Db 17526 SYTKSKSSSSSYKGRKRSYSCCTSRGAMSCRRCCYMRGASSMRGASSMRRAK 17585
|| ||| |
QY 915 TAAGCAGTGAACATCGCTTGTCTGACGGAACCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974
|| ||| |
Db 17586 RSWGGRSKWMTGGWRSKYIYCTGRRMMYIMCCCHRRRSYMYRSMRSMRSMRSMRSMRSMRSMRSMR 17645
|| ||| |
QY 975 CGTGATCTCTGACAGCGAGTGTCTGAACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1034
|| ||| |
Db 17646 GMRMGSASSRRCKSASRSRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMRSMR 17705
|| ||| |
QY 1035 CTTCCAGGCTGATCTTCTGCTGTAAAGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094
|| ||| |
Db 17706 SKRSSKYRGRGRKSMRTKSGSKGSKCWKRSSMTSSSCYIYSSASSCMMSKSCMC 17765
|| ||| |
QY 1095 GCTACCGCTGGAGGAACCGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1154
|| ||| |
Db 17766 CMMMRRCACCYSSNSCTSMYRCWGGMKSYSTGCTTCKCTGYKKSRTWYMYWYWS 17825
|| ||| |
QY 1155 AACCGTGTCTACCGGAGAACACCTC 1181
```

```

Db 17826 KMYWRAACMNYCYMSTRMMCMNSMC 17852
::: ::::::::::::::: I
ICYMANT      1249 bp    mRNA    linear    INV 18-SEP-1998
Ichthyophthirius multifiliis immobilization antigen precursor,
mRNA, partial cds.
M92907.1 GI:3628568
SOURCE       Ichthyophthirius multifiliis.
ORGANISM     Ichthyophthirius multifiliis.
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Ophryoglenina; Ichthyophthirius.
Lin.T.L. and Dickerson, H.W.
Purification and partial characterization of immobilization
antigens from Ichthyophthirius multifiliis
J. Protozool. 39 (4), 457-463 (1992)
93020590
MEDLINE      1383510
PUBMED       1383510
REFERENCE    2 (bases 1 to 1249)
AUTHORS     Clark, T.G., McGraw, R.A. and Dickerson, H.W.
TITLE       Developmental expression of surface antigen genes in the parasitic
            ciliate Ichthyophthirius multifiliis
PROC. NATL. ACAD. SCI. U.S.A. 89 (14), 6363-6367 (1992)
92335298
MEDLINE      1631132
PUBMED       1631132
REFERENCE    3 (bases 1 to 1249)
AUTHORS     Clark, T.
TITLE       Direct Submission
JOURNAL      Submitted (06-JUN-1992) Microbiology and Immunology, Cornell
            University, Ithaca, NY 14853, USA
4 (bases 1 to 1249)
DIRECT SUBMISSION
SUBMITTED (18-SEP-1998) Microbiology and Immunology, Cornell
University, Ithaca, NY 14853, USA
SEQUENCE UPDATE BY SUBMITTER
ON SEP 18, 1998 THIS SEQUENCE VERSION REPLACED GI:159289.
FEATURES
source
1..1249
/organism="Ichthyophthirius multifiliis"
/strain="Georgia"
/db_xref="taxon:5932"
/dev_stage="trophont"
<1..1190
/note="48 kDa"
/citation=[2]
/citation=[1]
/codon_start=3
/evidence=experimental
/transl_table=6
/product="immobilization antigen precursor"
/protein_id="AAC36158.1"
/db_xref="GI:3628569"
/translation="AVPCPDGQTQAGLTIDVGAADLGTCVNCRPNFYNGGAOGEAN
GNPFANNAARGICVPCCIINRVGSVTNAGDLATLCSTCCPTGTALDDGVTDVF
RSAQCCKPNFYNGSGSQGEAPGVFAGAAGAAVTSQCVPCQKNDSDPA
TACAQNLATQCSNQPTGTVLDGGTYLVFNFSATLCKVRPNFYNGSGSPQGEAPGV
QVFAAGAAQAANAQVTSQCVPCIINKNSPATGAQANLATOCSTQCPGTAIQDQGT
LVFNSNSTQSOCIANYFPNGLEAKSQCLKCPVSRTTPAHAPONTAOTACLITC
PAGTVLDGDTSTNFVASATECTKCSAGFFASKTTGTAGTDFTCTCKKLISGATAVN
IHQ"
<1..5
mat_peptide
slg_peptide
6..1187
/evidence=experimental
/product="immobilization antigen"
177..893
/citation=[2]
repeat_region
/rpt_type=tandem

```

```
/protein_id="AAL40941.1"
/db_xref="GI:1764245"
/translation="MSGGRFNFDDGSGYGGWEDKAHGHGVCCTPKGQGYTGSWS
HGFEVLGVYWPQSGVTQGTWAGQKRGHGLGLESKRWYKGEWTHGFKGRYGVRECA
NGAKYEGTWSNGLQDGYGTETSDG"
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN
Query Match 3.5%; Score 50; DB 9; Length 125020;
Best Local Similarity 10.1%; Pred. No. 0.0049;
Matches 93; Conservative 409; Mismatches 406; Indels 17; Gaps 3;
QY 274 AACGTGAAGTCTCTGCTGGAACCGCTATCGCTGGAGGAGCTACCGCTACGCTGCTATC 333
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17784 AGKSSRGGTGYMKKGGGKSSKWKGSSTSRRKSSAKSCSYMWGSMCMSCMSM 17725
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 334 ATCAACCGAGTGTGTAAGTGTGCGATCACTTACAAACGAGACGCTCTAACTTCAAC 393
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17724 AKSYMICYMYRMSYSSYYKYCYSCMGMSSTSYSCCMKMSWGSKCYCMKYYSWS 17665
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 394 GCTGGAGCTTCTACTGTACCGCTTCTGTGAACCGCTGGAGGAGCTCT-----GA 448
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17664 SYSTSGYTSSTCKKYKCSMSMYKCKTSYRKRYSYYWGGKRAKKIYICAGRRRR 17605
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 449 CCGCTGGAACCGCTCTACCATCGTGGCTCAGTGTAAAGTGTCTTCTACCGGAACCG 508
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17604 MSTWKCCAKWMSYCWYCYMYYSKCTYKSSCYKRGYYGSKTCYSAGGKRSRMY 17545
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 509 CTCTGACGACGAGTGACCCAGCTACTAGTGGCTCTTTTACCAGT-GTGTGAAGTGT 567
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17544 YCMRRSSKSSMSMARSSWCMGAGYRRSKRSAGWRRSKGRSTGMKRACSKK 17485
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 568 CGCCTGAACCTTCTACTAACGGAACAAACGGAACACCCCTTCAACCCCTGGAAGTCT 627
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17484 TGSYGTGRSMKKKGYSKYSGMKGKKTCTCMKYKYKRTSCWYMYKSWYKRYK 17425
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 628 CAGTGTACCCCTGTCTCTATCAAGCTGCTAAAGTGTGCTGAGCTACCGTACCGTGAAC 687
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17424 RCMKKKGCTGYRGMSSKSGYKSMRMRGSSYSTSCWKSCHGYSWMMKMYKSKR 17365
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 688 GAGCCTACCATCACCTCAGTGTAACTGCTGCTTCTCTGACGGAACCATCTCTCTGCT 747
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17364 SMRGSMSKSGWRYAGRGYSMSWSTRKRRSKCYSKYKGRKMGKMGWGMKSGKYW 17305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 748 GGAGTGAACAACGCTGCTGAGAACACCGAGTGTACCAAGTGTCTCTACTTCTAC 807
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17304 SSMKKMRSSWCKYSTKYSKGRSKGWRSTKSKSSMRMAGSKCTYSSYWSNNR 17245
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 808 AACAAACAGCTCCTAACTTCAACCTGGAACCTTACCTGTCTGCTTGTCTCTGCTAC 867
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17244 RNNR-----MGKTGCNTMYRRSRAMNGNAAAAGCTTCCCCANTNGGGGAAAA 17196
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 868 AAGACTACGAGCTAGGCTACCGCTGGAGGAGTGTCTACCTGTGCTAAGCAGTGTAAAC 927
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17195 AGGCGSASRASCYKGRMSKSCYRSGTRRCKMKCRGRSGRSGKSMGTRGRSGKTSYS 17136
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 928 ATCGCTTGTCTGACGGAACCGCTATCGCTTCTGAGCTACCAACTACGTGATCTCTGAG 987
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17135 AKSGRCCYCWGGRKGRKCMSSRMKMYKSYRRKRWMTCKMCKYSMYMAMYCRSMCC 17076
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 988 ACCGAGTGTCTGAAGTGTGCTTACTTCTACTTCTGACGGAACAACTTCCAGGCTGA 1047
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17075 MCMCKSCCGYSMGMSYSYSGYSMGMSYMRYSYRSTSKAWRSKRGMTGGRYK 17016
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 1048 TCTTCTCGCTGAAGGCTTCTCTGCTTAACAAGTGTGAGGAGCTGTGGCTACCGCTGA 1107
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 17015 GGRSYGKGGGYSYKGGSGWGGKSTGRSSAGKKSYYKMKMSCARYKSMYSKYCSRW 16956
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 1108 GGAACCCCTACCCCTGATCGCTCAGTGTGCTCTGAGAGTGTCTGCTGGAACCCGTGACC 1167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 16955 WCMYSYYCWGCKWNSCWMSMRKSKGWSWAKGMVMDKSVSTDKMBSRSBSKVKS 16896
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 1168 GACGGAACCACTTCTACTTACAGC 1192
```

```
Db 16895 KWSMSRRGKYYCTDSMKMKMYTGC 16871
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
RESULT 7
AC073693/c
LOCUS
DEFINITION
Mus musculus clone RP23-152L22, WORKING DRAFT SEQUENCE, 19
unordered pieces.
AC073693
AC073693.1 GI:8810310
HTGS_PHASE1; HTGS_DRAFT.
MUS MUSCULUS.
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 230372)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 230372)
Direct Submission
DOE Joint Genome Institute.
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1795526
Center clone name: RPCI-23_152L22
-----
Summary Statistics
Consensus quality: 217356 bases at least Q40
Consensus quality: 223517 bases at least Q30
Consensus quality: 224719 bases at least Q20
Estimated insert size: 258000; agarose-fp estimation
Estimated insert size: 228572; sum-of-contigs estimation
Quality coverage: 7.65 in Q20 bases; agarose-fp estimation
Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1087: contig of 1087 bp in length
* 1088 1187: gap of unknown length
* 1188 2715: contig of 1528 bp in length
* 2716 2815: gap of unknown length
* 2816 4158: contig of 1343 bp in length
* 4159 4258: gap of unknown length
* 4259 6473: contig of 2215 bp in length
* 6474 6573: gap of unknown length
* 6574 8982: contig of 2409 bp in length
* 8983 9083: gap of unknown length
* 9083 11286: contig of 2104 bp in length
* 11287 17000: contig of 5714 bp in length
* 17001 17101: gap of unknown length
* 17101 25179: contig of 8079 bp in length
* 25180 25279: gap of unknown length
* 25280 34875: contig of 9596 bp in length
* 34876 43664: contig of 8689 bp in length
* 43665 43764: gap of unknown length
* 43765 53788: contig of 10024 bp in length
* 53789 53888: gap of unknown length
* 53889 65520: contig of 11632 bp in length
```

```

* 65521 55620: gap of unknown length
* 80284 80284: contig of 14664 bp in length
* 80384 80384: gap of unknown length
* 98449 98449: contig of 18065 bp in length
* 98549 98549: gap of unknown length
* 98550 98550: contig of 17713 bp in length
* 116262 116262: contig of 17713 bp in length
* 116363 116363: gap of unknown length
* 140707 140706: contig of 24344 bp in length
* 140807 140806: gap of unknown length
* 167610 167609: contig of 26803 bp in length
* 167710 167709: gap of unknown length
* 199399 199398: contig of 31689 bp in length
* 199499 199498: gap of unknown length
* 199499 230372: contig of 30874 bp in length.
FEATURES
    source
        1. .230372
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-152L22"
            /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 61266 a 53740 c 52522 g 61042 t 1802 others
ORIGIN
Query Match          3.1%; Score 43.4; DB 2; Length 230372;
Best Local Similarity 53.9%; Pred. No. 0.46;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 166 GCTGCTGTTTGGTGGAGCTTCTACCTGTACCCCTTGTCTCAGAGAAGGAGCT 225
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212511 GCTGCTGCTCCCTGGCTGCTGCTCTCCCTGGCTGCTGCTGCTGCTGCT 212452

QY 226 GGAGCTACACCTAACCTCTGCTACCGCTAACCTGTGACCCAGTGTAAAGTGT 285
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212451 GCTGCTGCTCCCTGGCTGCTGCTACTGCTCCCTGGATGCTGCTCCCTGAATGGT 212392

QY 286 CTTGCTGGAAACCGCTATGCTGGAGGAGCTACCGACTACGCTGCT 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212391 GCTACTGCTCCCTGGCTGCTGCTGCTGCTGCTCCCTGGATGCTGCT 212347

RESULT 8
AC093866 159764 bp DNA linear PRI 01-MAR-2002
LOCUS Homo sapiens BAC clone RP11-582C12 from 4, complete sequence.
AC093866 AC025608
AC093866.3 GI:16973777
HTG.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 159764)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
2 (bases 1 to 159764)
Pearman, C., Meyer, R. and Doebber, A.
The sequence of Homo sapiens BAC clone RP11-582C12
Unpublished (2001)
3 (bases 1 to 159764)
Waterston, R.H.
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 159764)
Waterston, R.H.
Direct Submission
Submitted (17-NOV-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 17, 2001 this sequence version replaced gi:15887350.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0582C12
Drafting Center: WIBR
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-549C16, 2000 bp overlap;
the clone sequenced to the right is RP11-115D19. Actual start of
this clone is at base position 123633 of RP11-549C16; actual end is
at base position 159764 of RP11-582C12.

Sequence derived from one plasmid subclone, base position 55515 to
55572.

The sequence of AC025608 has been incorporated into AC093866.
FEATURES
    source
        1. .159764
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="4"
            /map="4"
            /clone="RP11-582C12"
            /clone_lib="RPCI-11"
            repeat_region 137..299
            repeat_region /rpt_family="ERV1"
            302..653
            /rpt_family="ERV1"

```

[illegible]


```
mRNA
join(<1..56,141..>643)
/product="sexually induced protein SIG 1"
CDS
join(<1..56,141..>643)
/codon_start=2
/product="sexually induced protein SIG 1"
/protein_id="AAK61135.1"
/db_xref="GI:14336350"
/translation="SHDMCTCDNRNFGAGCRLTCTPFGKAHVDTPRGDLASLTIGDH
YDIIILAGSTLVPYGTGTFPLMSDTACTVIANTAHDYNECSNKGLCDRIKLVCECPPG
YDGAACORASCPSKANSSTPGSGGSGRSNTNFKVFNANSAPFHGRAAEVVRDQCSCGHG
TCMTIEQLAFIDHNGTIDLMDKDTVM"
BASE COUNT    166 a   163 c   144 g   170 t
ORIGIN

Query Match
Best Local Similarity 55.2%; Pred. No. 1.4; Length 643;
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 278 TGAAGTCTCTGCTGAACCGGTATCGCTGGAGGACTACCGTACTGCTATCATCA 337
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 TCTCGGATACTGCTGGAACCGTTATCGCTAACACGGCTCATGACTACATGAATGTTCCA 341

QY 338 CCGAGTGTGTGAACGTGCGTCACACTTCTACACGAGCAAGCGTCCCTAACTTCAACGCTG 397
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 342 ATAAGGATTATCGCATCGCATCAATGCTTTTGGGAATGTCTCTCGGTATGACGGTG 401

QY 398 GAGCTTCTACCTGTATCCGGTTGTCC 422
||||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 CAGCTTGCACGGCTGCATCATGTCC 426

RESULT 10
AC099748
LOCUS
DEFINITION
AC099748
VERSION
AC099748.2 GI:21427692
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE
Bos taurus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 179217)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,S.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 179217)
Green,E.D.
Direct Submission
Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 179217)
Green,E.D.
Direct Submission
Submitted (15-JUN-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jun 15, 2002 this sequence version replaced gi:17017552.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov

----- Project Information
Center project name: crq
Center clone name: 504H16

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 177994 bases at least Q40
Consensus quality: 178622 bases at least Q30
Consensus quality: 178774 bases at least Q20
Insert size: 136000; agarose-fp
Insert size: 178817; sum-of-contigs
Quality coverage: 12.94x in Q20 bases; agarose-fp
Quality coverage: 9.85x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 32314: contig of 32314 bp in length
* 32315 32414: gap of unknown length
* 32415 66136: contig of 33722 bp in length
* 66137 66236: gap of unknown length
* 66237 80633: contig of 14397 bp in length
* 80634 80733: gap of unknown length
* 80734 92700: contig of 11967 bp in length
* 92701 92800: gap of unknown length
* 92801 179217: contig of 86417 bp in length.
* Location/Qualifiers
* 1..179217
* /organism="Bos taurus"
* /db_xref="taxon:9913"
* /clone="RP42-504H16"
* /clone_lib="RP42"
* 1..32314
* /notes="assembly_fragment
* clone_end:SP6
* vector_side:left"
* 32415..66136
* /notes="assembly_fragment"
* 66237..80633
* /notes="assembly_fragment"
* 80734..92700
* /notes="assembly_fragment"
* 92801..179217
* /notes="assembly_fragment
* clone_end:T7
* vector_side:right"
* 164910..179217
* /notes="clone overlaps with GenBank Accession Number
* AC113228 clone RP42-67823 (center project name crl)"
* BASE COUNT 53762 a 38501 c 36655 g 49899 t 400 others
* ORIGIN

Query Match 2.9%; Score 41; DB 2; Length 179217;
```


Thu Feb 20 11:09:56 2003

Db 326736 CCCGGAATTTCCGGATCGTTATACGTGATGCTGACAGCAATATACGGGAGAGTAACATT 326677
Qy 181 CTGTGAGCTTCTACCTGTACCCCTTCTCCTCAGAAAGAGGAGCGCTGGAGCTCAGCCTAAC 240
Db 326676 GCTGTAATTTGGTATCTGTACTCTGATCTGACTGATGCTGCTGCTGATTTCCCTTC 326617
Qy 241 CCTCTGCTACCCGCTAACCTGTGACCCAGTGTACGTGAAGTGTCTTCTGCTGGACCGCT 300
Db 326616 GGTATGCTGTGCTTAACCTGCTTAACGGTTATTTGGGGAGTATTCAGTTAGCGGA 326557
Qy 301 ATGCTGTGGAGGAGCTACCGCTACCTGCTATCATCACCGAGTGTGAACCTGTGCGCATC 360
Db 326556 TGTGAAGGTGAGAGTGGTTTTATCGGTGTAATTTGACTACCTGTGTTACAGGTGACGGT 326497
Qy 361 AACTTCTACAACGAGAACGCTCCTCACTTCAACGCTGGAGCTTCACTGTACCCGCTTGT 420
Db 326496 AACATCAGCTGTGGGTGCTCAGAACTAGGGTGTAACTGTGATGTTTACCACCTTC 326437
Qy 421 CCTGTGAACCGGTGGGAGGAGCTCTGACCGCTGGAACGC 461
Db 326436 AATGACTGTGCTGCCACCACTTTGAGTTATGGTAACGC 326396

RESULT 15
AC026803/c 224271 bp DNA linear PRI 16-JUL-2002
LOCUS AC026803 Homo sapiens chromosome 19 clone CTD-2639E6, complete sequence.
DEFINITION AC026803
ACCESSION AC026803
VERSION AC026803.7 GI:21844558
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 224271)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 224271)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 224271)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 224271)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 16, 2002 this sequence version replaced gi:21637469.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.3% of Sequence;
Estimated Total Number of Errors is 0.9.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 232kb). It is clipped at the overlap with AC009002. The
number of bases overlapped is 5260.
FEATURES
source 1..224271
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2639E6"
BASE COUNT 53830 a 58078 c 57960 g 54403 t
ORIGIN

Query Match 2.7%; Score 38.4; DB 9; Length 224271;
Best Local Similarity 49.5%; Pred. No. 14;
Matches 99; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
Qy 1056 CTGTAAGGCTTGTCTGCTTAACRAAGTGCAGGGAGCTGTGGCTACGCTGGAGGAACCGC 1115
Db 173706 CAGTAAGCCTTGGCGGTTTGTCTAAGGGCCAAAGCAANTGGCTTTCACGGTAATAGATGC 173647
Qy 1116 TACCTTGATCGCTCAGTGTGCTCTGGAGTGTCTCTGTAACCGTGTGCTGACCGACGGAAC 1175
Db 173646 TGCCAAGGAGATTTCAGATACATCTCTGCAAAATGCACCTGACCCGCTCGCGCACTGGCAAC 173587
Qy 1176 CACCTCTACCTTACAGAGGCTGCTTCTGTAGTGTGTGAAGTGTGCTGCTTAACCTTCTACAC 1235
Db 173586 CATATTTAATGGCAAGCACTTTTACGCTGGATATTGGAAGGGCTCCTCTGCCCTCACCTC 173527
Qy 1236 CACCAAGCAGACCGACCTGGG 1255
Db 173526 CATAAGGAAGAGGCTCGGG 173507
Search completed: February 16, 2003, 22:05:16
Job time : 3318.14 secs